J Comput Neurosci (2014) 37:29–48
DOI 10.1007/s10827-013-0486-0

Encoding certainty in bump attractors

Samuel R. Carroll · Krešimir Josić · Zachary P. Kilpatrick

Received: 23 August 2013 / Revised: 24 October 2013 / Accepted: 29 October 2013 / Published online: 24 November 2013

Abstract Persistent activity in neuronal populations has been shown to represent the spatial position of remembered stimuli. Networks that support bump attractors are often used to model such persistent activity. Such models usually exhibit translational symmetry. Thus activity bumps are neutrally stable, and perturbations in position do not decay away. We extend previous work on bump attractors by constructing model networks capable of encoding the

produce reliable bumps is essential for understanding the mechanism behind spatial working memory.

Models that can store a continuous range of spatial locations typically possess solutions that are neutrally stable (Amari 1977; Seung 1996; Brody et al. 2003). Due to the neutral stability of such attractors, perturbations that change the location of a bump of activity do not decay away (Amari 1977; Camperi and Wang 1998; Compte et al. 2000). Aside from experimentally-introduced distractors in spatial working memory experiments (Miller et al. 1996), cue memories can also be degraded by internal variability within cortical networks (Faisal et al. 2008). Stochastic models show that such variability causes bump attractors to wander diffusively, due to their inherent neutral stability (Camperi and Wang 1998; Compte et al. 2000; Laing and Chow 2001; Kilpatrick and Ermentrout 2013). Psychophysical studies show that errors made recalling remembered spatial locations scale roughly linearly with delay time, suggesting the remembered location may diffuse in time (White et al. 1994; Ploner et al. 1998). Also, heterogeneities in the spatial structure of the underlying neuronal network can further degrade the relation between the stored memory and the initial cue (Seung 1996; Renart et al. 2003; Itskov et al. 2011; Hansel and Mato 2013). One solution to this problem is to structure the spatial arrangement of excitatory synapses (Kilpatrick and Ermentrout 2013; Kilpatrick et al. 2013) to make networks robust to dynamic and static parametric perturbations. Thus, the spatial organization of synaptic architecture can play a major role in accurately encoding stimuli for future recall.

To explore the relation between network architecture and the neural computation underlying working memory, we consider bump attractor networks capable of encoding cue *certainty*. We define *certainty* as the likelihood that the presented cue was faithfully communicated to the network generating delay period activity. A number of experimenwtevelTJ0TcETBT10001051.999276.415Tm/Tc11Tf[c/Cs(t)-2(od-6(d)2(e)-2.2(n)1

2 Bump attractor networks

Bump attractor networks were originally developed as general models of recurrent neuronal circuits that can support spatiotemporal patterns of activity (Wilson and Cowan 1973; Amari 1977). Since then they have been used to represent activity subserving spatial working memory (Camperi and Wang 1998; Compte et al. 2000) and visual orientation processing (Ben-Yishai et al. 1995). We consider a spatially organized neural field model where the positions of neurons correspond to their preferred stimulus orientation. We focus on a ring architecture, but we believe these ideas will extend to more general models.

Consider a single population which incorporates local excitation and broadly tuned inhibition (Amari 1977; Ben-Yishai et al. 1995; Ermentrout 1998)

$$\frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_{-\pi}^{\pi} w(x-y) \times f(u(y,t))dy + I(x,t). \quad (2.1)$$

Here $u(x,t)$

We note that in the limit of fast inhibition, τ

Fig. 1 Stationary bump solutions. **a** *Blue* and *black* indicate neutrally stable solutions for $U(x)$ while *red* indicates unstable solutions that are attracted to the boundary of

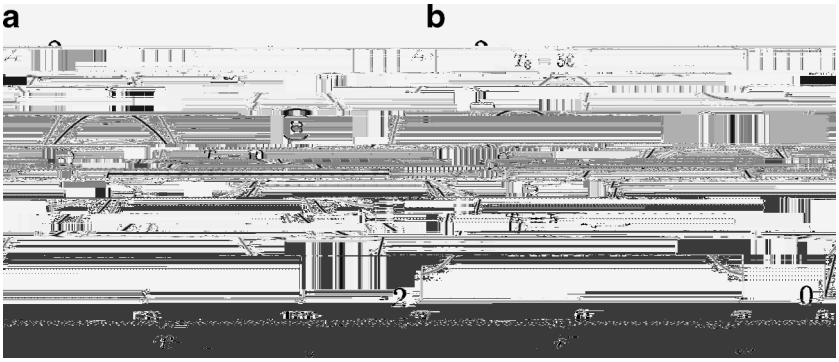

Fig. 2 The integration of constant external input $I(t) = I_0$ lasting for T_0 time units by the single population network Eq. (2.1). **a** The amplitude of the bump in response to a current injection of the form Eq. (3.7). **b** Bump solution as $t \to \infty$

where, for convenience, we use the normalized stochastic variable $\xi_A(t) = \xi(t)/A_0$

Note that since $\mathcal{L}^{\square}\phi_e(x) = \mathcal{L}^{\square}\phi_o(x) = 0$, we use Eq. (3.14) and the general weight function given by Eq. (2.8) to find that

$$\phi_o(x) = f^{\square}(U(x)\sum_{k=}^{N}$$

Fig. 5 Amplitudes of the stationary bump solutions $U(x) = A_0 + A_1 \cos x$ for varying $\bar{w}$ values as per equation Eq. (4.2). **a** A_0 as a function of $\bar{w}$. **b** A_1 as a function of $\bar{w}$. Other values used: $s = \frac{2}{\pi}$, $\bar{w}_{ee} = 1$

we conclude that one of the major differences between networks with one and two populations is that it is possible to destabilize stationary bumps with sufficiently slow inhi-

field models. In particular, we examine the stability of stationary bump solutions of the form $U(x) = A\cos x$ when the firing rate function has the form given in Eq. (2.3). Hence,

$$(\lambda + 1)\mathcal{A}_1 = \begin{cases} \mathcal{A}_1 s w_1 \left[\frac{\pi}{2} - \cos^{-1}\left(\frac{1}{sA}\right) - \frac{1}{sA}\sqrt{1 - \left(\frac{1}{sA}\right)^2} \right], & \text{for } sA > 1, \\ \mathcal{A}_1 s w_1 \frac{\pi}{2}, & \text{for } sA \le 1, \end{cases} \tag{A.7}$$

$$(\lambda + 1)\mathcal{B}_1 = \begin{cases} \mathcal{B}_1 s w_1 \left[\frac{\pi}{2} - \cos^{-1}\left(\frac{1}{sA}\right) + \frac{1}{sA}\sqrt{1 - \left(\frac{1}{sA}\right)^2} \right], & \text{for } sA > 1, \\ \mathcal{B}_1 s w_1 \frac{\pi}{2}, & \text{for } sA \le 1, \end{cases} \tag{A.8}$$

and $\mathcal{A}_k = \mathcal{B}_k = 0$ for $k \neq 1$ and $\mathcal{A}_0 = 0$. Therefore, only $\sin x$ and $\cos x$ are eigenfunctions of the linearized system. All other Fourier modes cos

and, for the v equation

$$V(x) = \bar{w}_{ei} \int_{-\pi}^{\pi} (1 + \cos(x - y)) f(U(y)) dy,$$

so that

$$M_0 = \begin{cases} 2s\bar{w}_{ei} A \left[1 - \sqrt{1 - \left(\frac{1}{sA}\right)^2} \right] \\ \qquad + 2\bar{w}_{ei} \cos^{-1}\left(\frac{1}{sA}\right), & \text{for } sA > 1, \\ 2s\bar{w}_{ei} A, & \text{for } sA \le 1, \end{cases}$$

$$M_1 = \begin{cases} sA\bar{w}_{ei} \left[\frac{\pi}{2} - \cos^{-1}\left(\frac{1}{sA}\right) - \frac{1}{sA}\sqrt{1 - \left(\frac{1}{sA}\right)^2} \right] \\ \qquad + 2\bar{w}_{ei}\sqrt{1 - \left(\frac{1}{sA}\right)^2}, & \text{for } sA > 1, \\ sA\bar{w}_{ei}\frac{\pi}{2}, & \text{for } sA \le 1. \end{cases} \tag{B.5}$$

Again, we have a cont

Eq. (B.9), we see that when using the weight functions in Eq. (2.5) we have the system

$$(\lambda + 1)\mathcal{A}_0 = \bar{w}_{ee} \int_{-\pi}^{\pi} (\mathcal{A}_0 + \mathcal{A}_1 \cos y + \mathcal{B}_1 \sin y) f'(U(y)) dy$$
$$- \bar{w}_{ie} \int_{-\pi}^{\pi} (\quad_0 + \quad_1 \cos y + \quad_1 \sin y) dy,$$

$$(\lambda + 1)$$

Vijayraghavan, S., Wang, M., Birnbaum, S.G., Williams, G.V., Arnsten, A.F.T. (2007). Inverted-u dopamine d1 receptor actions